



**ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL**  
**Facultad de Ingeniería en Electricidad y Computación**

# **SISTEMA PARA EL ANÁLISIS Y PROCESAMIENTO DE LOS LOGS DE LOS SERVIDORES DE RED DE LA FACULTAD DE INGENIERÍA EN ELECTRICIDAD Y COMPUTACIÓN (FIEC) DE LA ESPOL USANDO HADOOP**

## **INTEGRANTES:**

Eddy Roberto Espinosa Daquilema  
Josué Jefferson Guartatanga Robayo

# PROBLEMA A RESOLVER

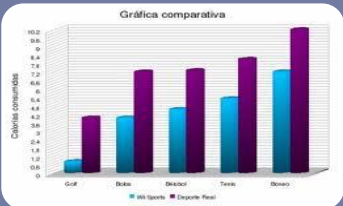
## Datos vs. Información



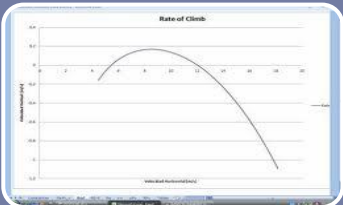
# JUSTIFICACIÓN



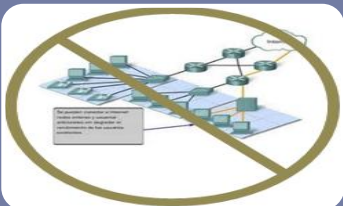
Existen herramientas con precios altos



No muestran resultados deseados



Rendimiento deficiente / tiempo extenso



No son distribuidas ni escalables

# ALCANCE

Información de los servidores: Cedro(HTTP), Ceibo(Maillog) y Palma(Samba).



Visitas/Recursos/Navegadores



Correos



Acceso de usuarios/Recursos



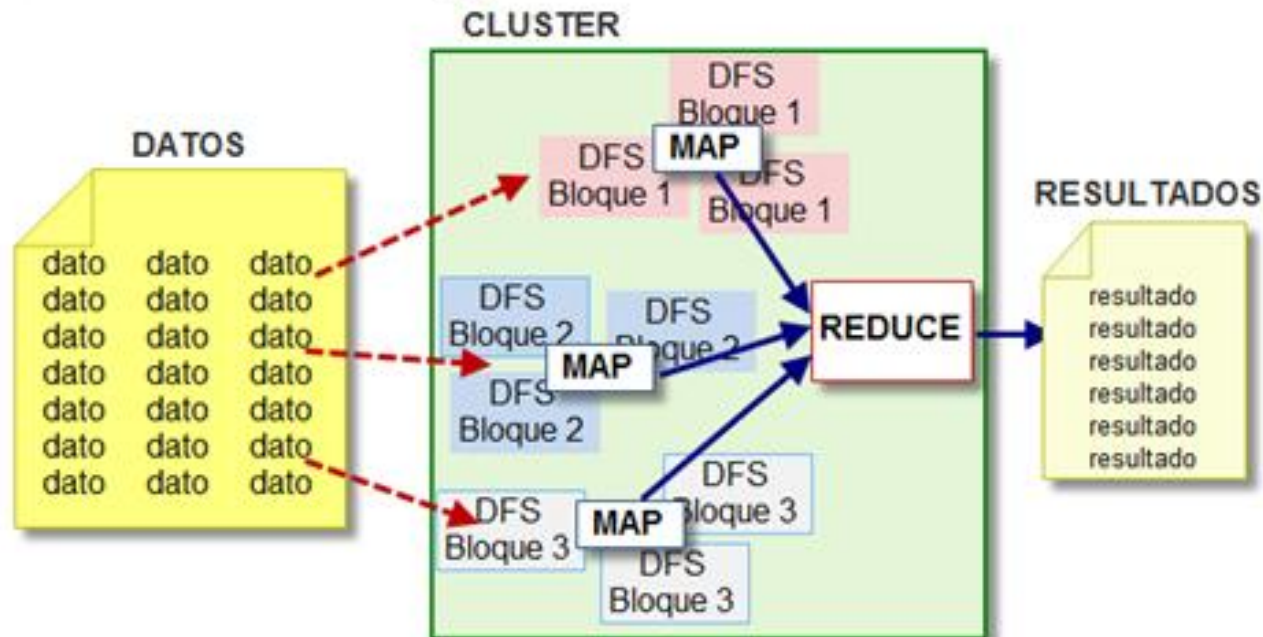
# HADOOP

Plataforma -> aplicaciones escalables

Desarrollador enfoca -> lógica de negocio

Corre en cluster / (HDFS)

Map/Reduce

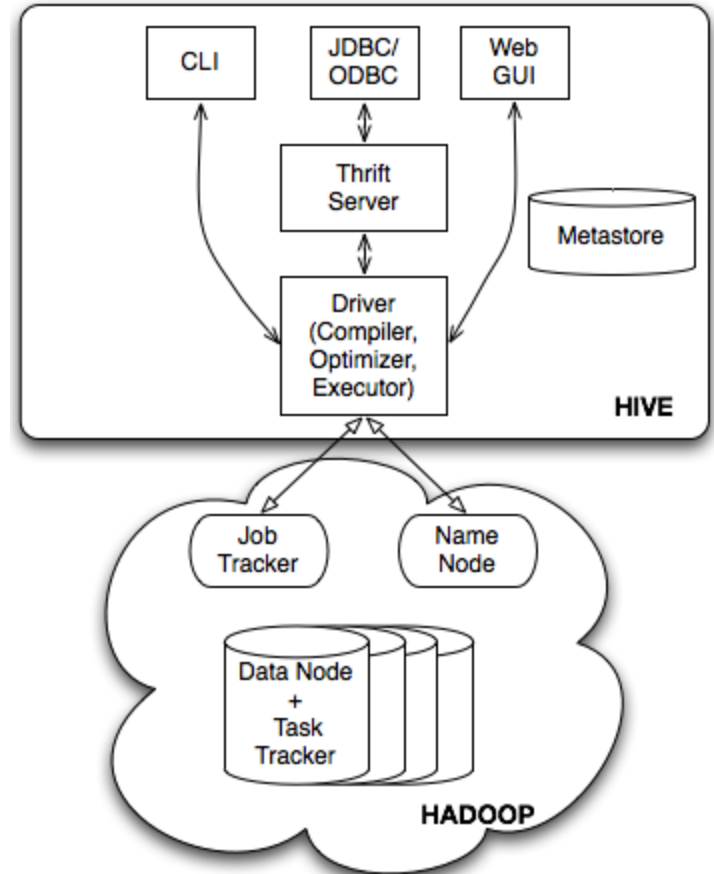


# HIVE

Infraestructura para data warehouse que provee sumariación de datos y soporta Ad-hoc queries

## Aplicaciones

- Logs
- Minería Datos
- Indexación documentos
- Inteligencia Negocios
- Modelamiento predictivo y Prueba de hipótesis



# FORMATO DE LOGS

## Cedro posee un Servidor Apache HTTP

```
67.195.112.238 - - [22/Nov/2009:04:13:29 -0500] "GET /resources/materias/licred/FIEC06221_implementacion_soporte_xp_prof.pdf HTTP/1.0" 200 51590 "-" "Mozilla/5.0 (compatible; Yahoo! Slurp/3.0; HTTP://help.yahoo.com/help/us/ysearch/slurp)"
```

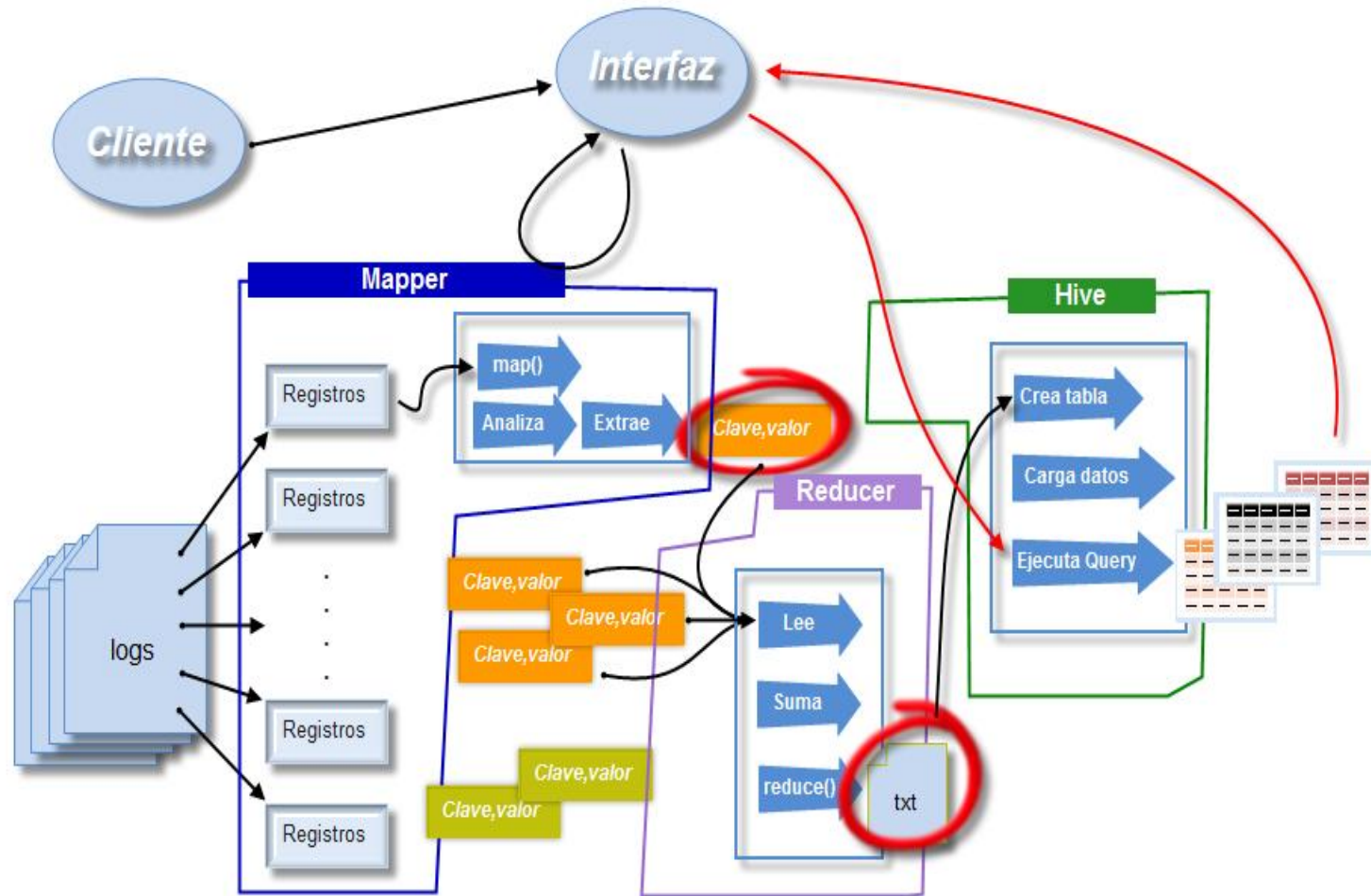
## Ceibo genera archivos tipo Maillog

```
Nov 9 00:21:16 ceibo milter-greylist: nA95LGs1020166: addr mail.periodicourbano.com[200.125.135.132] from <nobody@mail.periodicourbano.com> to <amera@fiiec.espol.edu.ec> delayed for 00:05:00 (ACL 300)
```

## Palma posee un servidor Samba.

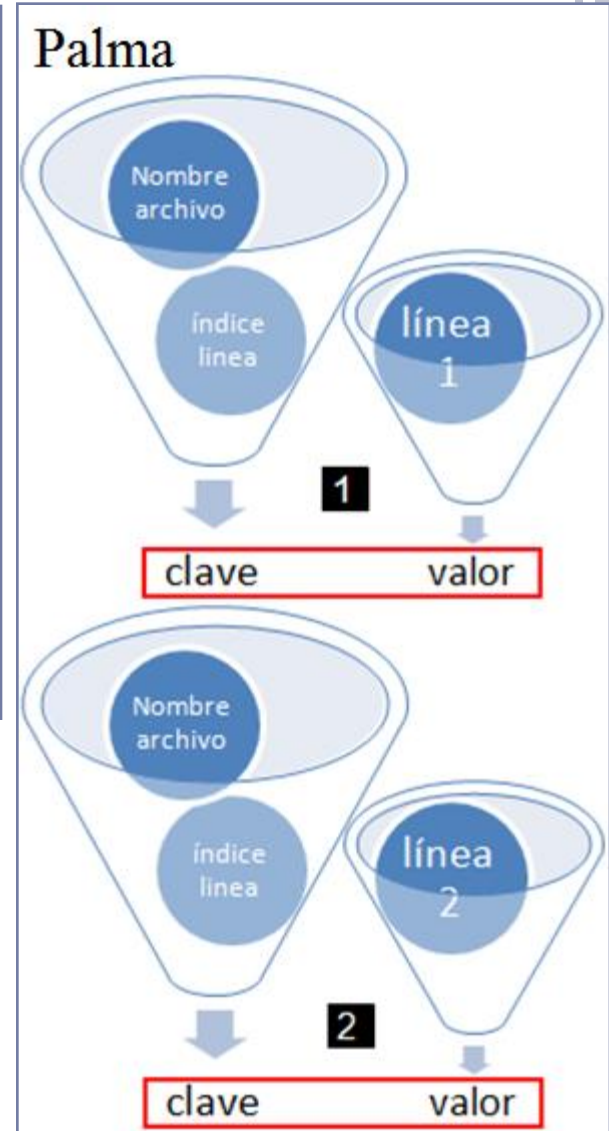
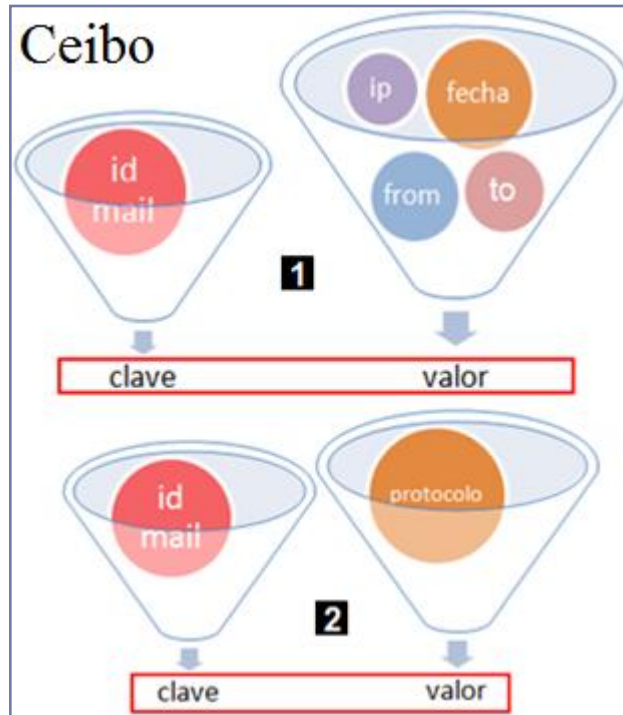
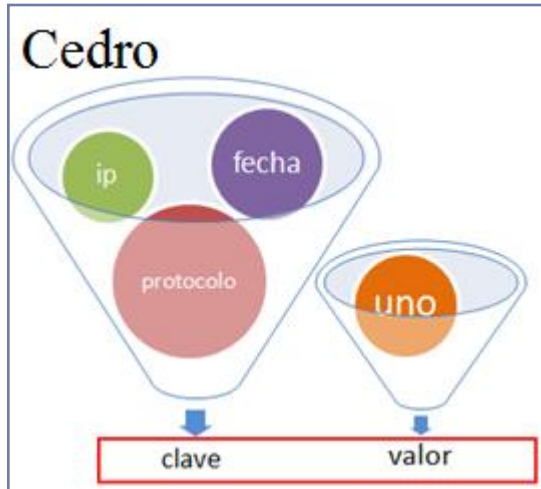
```
[2009/09/16 12:46:56, 1] smbd/service.c:make_connection_snum(1033) wrks126-170fiiec (200.9.176.170) connect to service apincay initially as user apincay (uid=6828, gid=501) (pid 5491) [2009/09/17 13:52:46, 2] smbd/open.c:open_file(391) opreciad opened file Internet Download Manager/IDMan.exe read=Yes write=No (numopen=11)
```

# DISEÑO DE LA SOLUCIÓN

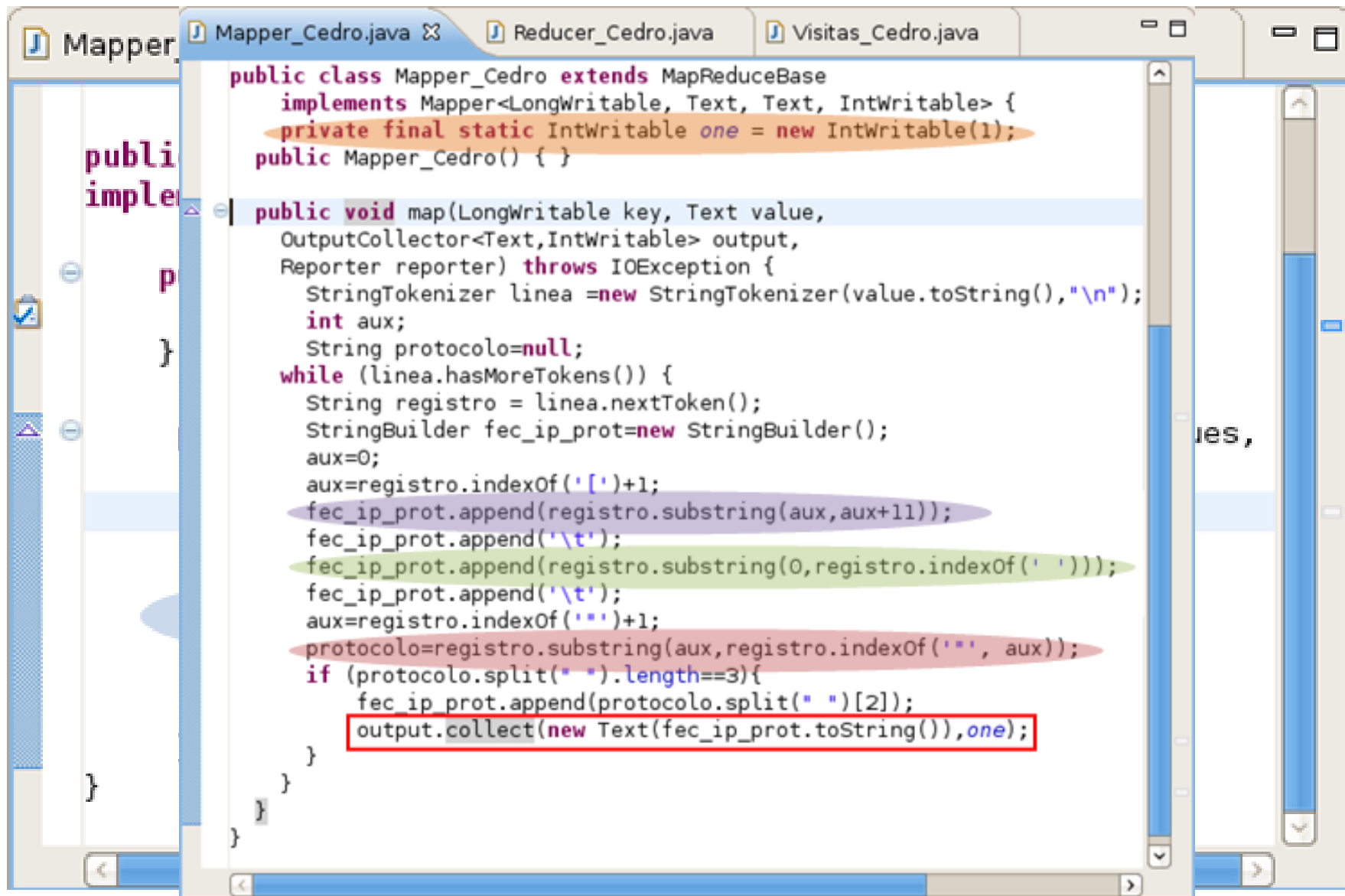




# CLAVES Y VALORES



# CÓDIGO BASE: MAPPER/REDUCER DE CEDRO



```
public class Mapper_Cedro extends MapReduceBase
    implements Mapper<LongWritable, Text, Text, IntWritable> {
    private final static IntWritable one = new IntWritable(1);
    public Mapper_Cedro() { }

    public void map(LongWritable key, Text value,
        OutputCollector<Text, IntWritable> output,
        Reporter reporter) throws IOException {
        StringTokenizer linea = new StringTokenizer(value.toString(), "\n");
        int aux;
        String protocolo = null;
        while (linea.hasMoreTokens()) {
            String registro = linea.nextToken();
            StringBuilder fec_ip_prot = new StringBuilder();
            aux = 0;
            aux = registro.indexOf('(') + 1;
            fec_ip_prot.append(registro.substring(aux, aux + 11));
            fec_ip_prot.append('\t');
            fec_ip_prot.append(registro.substring(0, registro.indexOf(' ')));
            fec_ip_prot.append('\t');
            aux = registro.indexOf('"') + 1;
            protocolo = registro.substring(aux, registro.indexOf('"', aux));
            if (protocolo.split(" ").length == 3) {
                fec_ip_prot.append(protocolo.split(" ")[2]);
                output.collect(new Text(fec_ip_prot.toString()), one);
            }
        }
    }
}
```

# CÓDIGO BASE: MAPPER/REDUCER DE CEIBO

```
Mapper_Ceibo.java 33  Reducer_Ceibo.java  Correos_Ceibo.java

identificador=registro.indexOf("milter-greylist: ");
if(identificador!=-1){
    identificadormail=new Text(registro.substring(identificador+17,identificador+31));
    int indice=registro.indexOf(": addr ");
    if (indice!=-1){
        fecha_ip_from_to.append(registro.substring(0,6));
        fecha_ip_from_to.append("\t");
        indice=registro.indexOf('[',indice+7);
        if(indice!=-1){
            fecha_ip_from_to.append(registro.substring(indice+1, registro.indexOf(']',indice+1)));
            fecha_ip_from_to.append("\t");
            indice=registro.indexOf(" from <");
            if (indice!=-1){
                fecha_ip_from_to.append(registro.substring(indice+7,registro.indexOf('>',+indice)));
                fecha_ip_from_to.append("\t");
                indice=registro.indexOf(" to <");
                fecha_ip_from_to.append(registro.substring(indice+5,registro.indexOf('>',+indice)));
                output.collect(identificadormail,new Text(fecha_ip_from_to.toString())); 1
            }
        }
    }
}
else{
    identificador= registro.indexOf("sendmail(");
    if(identificador!=-1){
        int indice= registro.indexOf("]: ", identificador);
        if(indice!=-1){
            identificadormail=new Text(registro.substring(indice+3, indice+17));
            indiceprotocolo=registro.indexOf(", proto=", indice);
            if(indiceprotocolo!=-1){
                Protocolo=registro.substring(indiceprotocolo+8,registro.indexOf(", ",indiceprotocolo+9));
                output.collect(identificadormail,new Text(Protocolo)); 2
            }
        }
    }
}
```

# CÓDIGO BASE: MAPPER/REDUCER DE PALMA

```
public class Mapper_Palma extends MapReduceBase
    implements Mapper<LongWritable, Text, Text, Text> {
    public Mapper_Palma() { }
    public void map(LongWritable key, Text value, OutputCollector<Text,Text> output,
        Reporter reporter) throws IOException {
        FileSplit fileSplit=(FileSplit)reporter.getInputSplit();
        String fileName= fileSplit.getPath().getName();
        StringBuilder llave=new StringBuilder();
        int index=0;
        String indexstring;
        * if(value.toString().charAt(0)=='['){
            index=value.toString().length()+1+Integer.parseInt(key.toString());
            indexstring= String.valueOf(index);
            llave.append(fileName);
            llave.append(" ");
            llave.append(indexstring);
            output.collect(new Text(llave.toString()),value); 1
        }
        else{
            index=Integer.parseInt(key.toString());
            indexstring= String.valueOf(index);
            llave.append(fileName);
            llave.append(" ");
            llave.append(indexstring);
            output.collect(new Text(llave.toString()),value); 2
        }
    }
}
```

# TIEMPOS

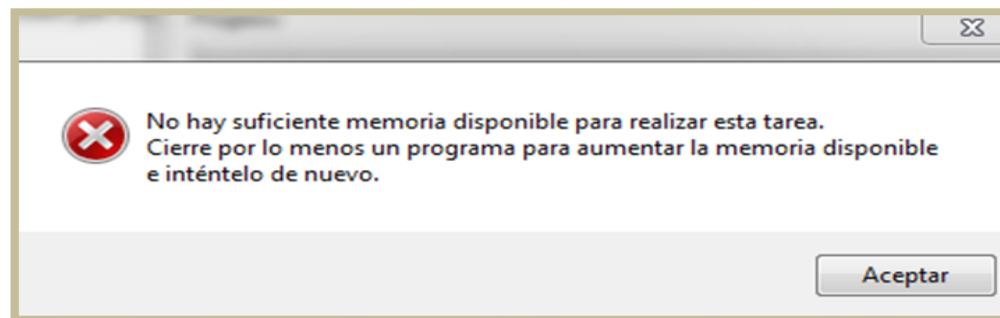
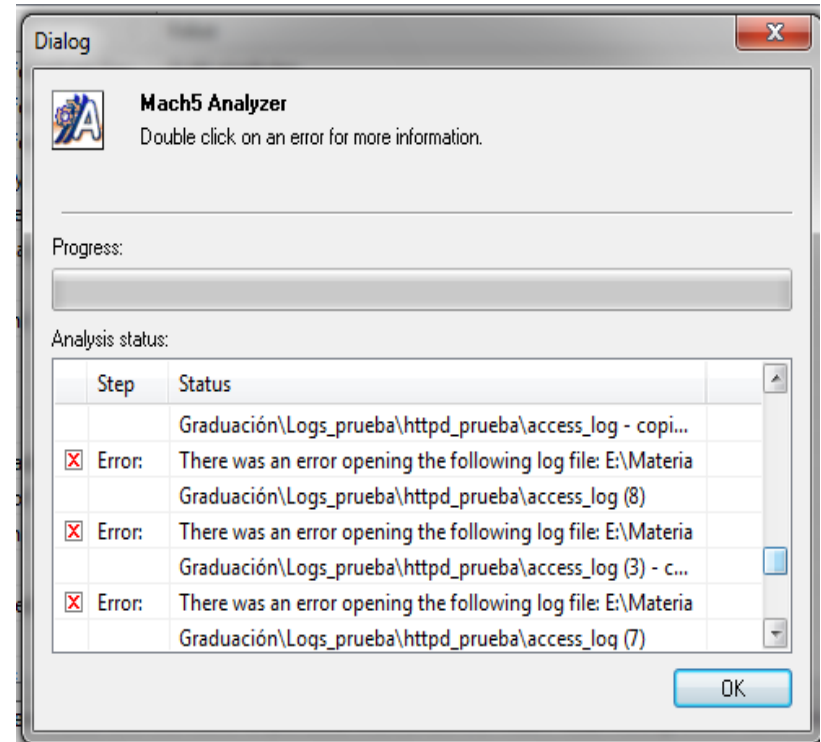
Tiempo (minutos)

		tiempo (min)								
		carga 37.2 MB / 0.036 GB								
		Palma Concatena			Palma Accesos			Palma Recursos		
		Nodos								
# prueba		5	10	15	5	10	15	5	10	15
	1	6,783	1,933	1,367	0,467	0,350	0,267	0,417	0,620	0,333
	2	6,833	1,967	1,983	0,450	0,620	0,317	0,533	0,220	0,350
	3	7,017	3,020	1,683	0,517	0,200	0,300	0,550	0,240	0,350
	4	7,683	3,517	1,683	0,550	0,210	0,333	0,533	0,220	0,350
	5	6,833	4,350	1,267	0,517	0,270	0,300	0,433	0,600	0,300
	6	6,383	4,417	2,050	0,717	0,220	0,317	0,517	0,220	0,317
	7	9,683	5,667	1,633	0,483	0,210	0,300	0,450	0,630	0,400
	8	6,167	4,317	1,483	0,467	0,600	0,283	0,450	0,240	0,300
	9	8,633	4,267	1,283	0,433	0,590	0,417	0,500	0,680	0,317
	10	6,350	5,367	1,333	0,517	0,230	0,317	0,517	0,200	0,317
	0	promedio	7,237	3,882	1,577	0,512	0,350	0,315	0,490	0,387
	Var. Std.	1,124	1,273	0,281	0,081	0,180	0,040	0,048	0,213	0,030

# APLICACIÓN NO DISTRIBUIDA

## Carga

- 10 GB





# APLICACIÓN DISTRIBUIDA

File Help

Parámetros de ejecución y consulta

Aplicativo:

Visitas a Cedro

▼

Filtro

Correos:

Remitente

▼

☐ Protocolo:

HTTP

▼

☐ Mes:

Enero

▼

Items:

10

▲▼

Ejecutar

Consultar

Salir

Detalles de Visitas a Cedro

Item	Fecha	Protocolo	I.P.	Frecuencia

Parámetros de ejecución y consulta

Aplicativo:

Visitas a Cedro

▼

Filtro

Correos:

Remitente

▼

☒ Protocolo:

HTTP

▼

☒ Mes:

Septiembre

▼

Items:

10

▲▼

Ejecutar

Consultar

Salir

Visitas a Cedro

▼

Visitas a Cedro

Recursos de Cedro

Navegadores de Cedro

Correos de Ceibo

Accesos a Palma

Recursos de Palma

HTTP

▼

HTTP

HTTPS

HFS

SSH

ICMP

DNS

IMAP

SMTP

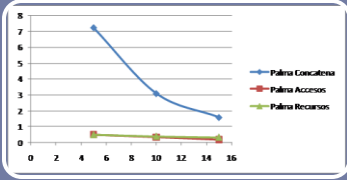
Remitente

▼

Remitente

Destinatario

# CONCLUSIONES



Estable entre 5 y 10 nodos con carga 1GB.



Procesamiento secuencial es extenso con archivos de gran tamaño.



El modelo distribuido es una solución escalable y la tolerancia a fallos se puede controlar.





## RECOMENDACIONES



Asignar suficiente espacio en disco para la instalación.



Familiarizarse con el entorno que ofrece Hadoop y su aplicación Hive.



FIN

**MUCHAS GRACIAS !!!**

